LITERATURE SURVEY

Date	15 March 2023
Team ID	NM2023TMID18999
Project Name	Optimizing Spam Filtering with Machine Learning

1. Temidayo et al says that "This study evaluated and compared the performance of two ensemble   models based on the random forest and extreme gradient boost ensemble algorithms. Baseline random forest and XGBoost spam detection models were developed based on the train/test split technique using the default parameters".

2. Mendizabal et al says that "This study has introduced the formulation of three different dimensionality reduction strategies to use when texts are represented by using synsets (an earlier lossless one called SDRS, a new low-loss one and a lossy one). The strategies were defined as optimization problems".

3. Thomas merly et al says "In order to filter the spam information in social media, a Chimp Sailfish Optimization-based Deep Neuro Fuzzy Network (ChSO-based DNFN) is proposed. The proposed method effectively performs well under high dimensional data in real platform environment using deep learning classifier. It is more robust and generates optimal result and also reduces the computational complexity problems"

4. Berrou et al says that "the current paper presents an intelligent and automated solution to spam emails detection using the logistics regression model trained by a teaching, learning, based optimization algorithms. Keywords: Logistic regressions. TLBO algorithm, Spam email detection.

5. Srinivasa Rao et al says that "to optimize the parameters of the network and to improve the accuracy, the optimization algorithm Rat Swarm Optimization (RSO) is used. Then, AFINN and SentiWordNet are used for sentiment analysis. This framework is evaluated on the three benchmark datasets; when comparing the performance of proposed method on the three dataset, spam assassin dataset achieves better spam detection accuracy of 99.82%".

6. Ghosh et al says that, "In terms of accuracy, the Random Forest classifier performs best and the performance of the Naïve Bayes classifier is substandard compared to the rest of the classifiers. Random Forest classifier had the accuracy of 99.91% and 99.93% for the Spam

Corpus and Spambase datasets respectively. The naïve Bayes classifier had the accuracy of 87.63% and 79.53% for the Spam Corpus and Spambase datasets respectively".

7. Rustam et al says that, "Although several spam detection approaches exist, detecting new strains of spam messages is challenging that requires a reliable and efficient intelligent spam email detection approach. This study utilizes features from the text of emails to determine whether it is spam or normal. Multiple features are combined to obtain a higher accuracy for spam email detection. Experiments involve machine learning and deep learning models and the influence of data resampling is also investigated".

8. Y Li et al says that, "This article presents an investigation of how machine-learning-based algorithms are used in email spam filtering by providing some previous researches that have shown to be successful. Spamming has raised security concerns as it causes potential loss to the users with fake advertisements, invalid information, undetected virus, and other harmful information".

9. Zivkovic et al says that "The proposed approach was validated on a public benchmark dataset (CSDMC2010) and compared to other cutting-edge techniques. The obtained results indicate that the suggested hybrid approach outperforms other spam detection solutions included in the comparative analysis, by achieving the highest classification accuracy".

10. Salama et al says that the, "transfer learning and data augmentation are employed to address the issue of a shortage of labeled data. In our implementation, the fully connected (FC) layer in the aforementioned pre-trained models is replaced with a Support Vector Machine (SVM) classifier, resulting in an improved accuracy".

REFERENCES:

1. Omotehinwa, Temidayo Oluwatosin, and David Opeoluwa Oyewola. "Hyperparameter Optimization of Ensemble Models for Spam Email Detection." *Applied Sciences* 13.3 (2023): 1971.

2. de Mendizabal, Iñaki Vélez, et al. "Multi-objective evolutionary optimization for dimensionality reduction of texts represented by synsets." *PeerJ Computer Science* 9 (2023): e1240.

3. Thomas, Merly, and B. B. Meshram. "ChSO-DNFNet: Spam detection in Twitter using feature fusion and optimized Deep Neuro Fuzzy Network." *Advances in Engineering Software* 175 (2023): 103333
.
4. Berrou, B. Khadija, et al. "Training a Logistic Regression Machine Learning Model for Spam Email Detection Using the Teaching-Learning-Based-Optimization Algorithm." *Proceedings of the 1st International Conference on Innovation in Information Technology and Business (ICIITB 2022).* Vol. 104. Springer Nature, 2023
.
5. Srinivasarao, Ulligaddala, and Aakanksha Sharaff. "Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages." *Multimedia Tools and Applications* (2023): 1-31.

6. Ghosh, Argha, and A. Senthilrajan. "Comparison of machine learning techniques for spam detection." *Multimedia Tools and Applications* (2023): 1-28.

7. Rustam, Furqan, et al. "Detecting ham and spam emails using feature union and supervised machine learning models." *Multimedia Tools and Applications* (2023): 1-17.

8. Li, Yixuan. "Algorithm Comparison on Email Spam Filtering Task." *Highlights in Science, Engineering and Technology* 34 (2023): 143-148.

9. Zivkovic, Miodrag, et al. "Training Logistic Regression Model by Hybridized Multi-verse Optimizer for Spam Email Classification." *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 2*. Singapore: Springer Nature Singapore, 2023.

10. Salama, Wessam M., Moustafa H. Aly, and Yasmine Abouelseoud. "Deep learning-based spam image filtering." *Alexandria Engineering Journal* 68 (2023): 461-468